

## RESEARCH ARTICLE

## Open Access



# Coevolution of paired receptors in *Xenopus* carcinoembryonic antigen-related cell adhesion molecule families suggests appropriation as pathogen receptors

Wolfgang Zimmermann<sup>1,2\*</sup>  and Robert Kammerer<sup>3</sup>

## Abstract

**Background:** In mammals, CEACAM1 and closely related members represent paired receptors with similar extracellular ligand-binding regions and cytoplasmic domains with opposing functions. Human CEACAM1 and CEACAM3 which have inhibitory ITIM/ITSM and activating ITAM-like motifs, respectively, in their cytoplasmic regions are such paired receptors. Various bacterial pathogens bind to CEACAM1 on epithelial and immune cells facilitating both entry into the host and down-regulation of the immune response whereas interaction with granulocyte-specific CEACAM3 leads to their uptake and destruction. It is unclear whether paired CEACAM receptors also exist in other vertebrate clades.

**Results:** We identified more than 80 *ceacam* genes in *Xenopus tropicalis* and *X. laevis*. They consist of two subgroups containing one or two putative paired receptor pairs each. Analysis of genomic sequences of paired receptors provide evidence that their highly similar ligand binding domains were adjusted by recent gene conversion events. In contrast, selection for diversification is observed among inhibitory receptor orthologs of the two frogs which split some 60 million years ago. The allotetraploid *X. laevis* arose later by hybridization of two closely related species. Interestingly, despite the conservation of the genomic landscape surrounding the homeologous *ceacam* loci only one locus resembles the one found in *X. tropicalis*. From the second *X. laevis* locus more than 80 % of the *ceacam* genes were lost including 5 of the 6 paired receptor genes. This suggests that once the gene for one of the paired receptors is lost the remaining gene cluster degrades rapidly probably due to lack of selection pressure exerted by pathogens.

**Conclusions:** The presence of paired receptors and selection for diversification suggests that also in amphibians CEACAM1-related inhibitory proteins are or were used as pathogen receptors.

**Keywords:** Carcinoembryonic antigen-related cell-cell adhesion molecule (CEACAM), Paired receptors, Positive selection, Allotetraploid, *Xenopus*, Amphibia, ITIM, ITAM, Immunoglobulin superfamily

## Background

A number of families of cell surface receptor with very similar extracellular domains and inhibitory or activating intracellular signaling motifs have been identified in vertebrates. The best investigated families represent the KIR, Ly49, Nkpr, SIGLEC, SIRP and CEACAM families [1]. These so called paired receptors are commonly encoded

in the same gene cluster and some are thought to play a role in homeostasis of the immune system by controlling activation and downregulation of immune reactions [2]. Many of the inhibitory members of paired receptors are expressed on natural killer (NK) cells where they sense major histocompatibility antigens (MHC) present on uninfected cells leading to tolerance. Loss of MHC expression frequently found in virus infected cells releases NK cell inhibition with concomitant destruction of the infected cells. Activating members of paired receptors seem to have evolved to counter common virus immune escape mechanisms in serving as decoy receptors on NK cells.

\* Correspondence: [wolfgang.zimmermann@med.uni-muenchen.de](mailto:wolfgang.zimmermann@med.uni-muenchen.de)

<sup>1</sup>Tumor Immunology Laboratory, LIFE Center, University Clinic, Ludwig-Maximilians-University, Feodor-Lynen-Str. 19, 81377 Munich, Germany

<sup>2</sup>Department of Urology, University Clinic, Ludwig-Maximilians-University, Marchioninistr. 15, 81377 Munich, Germany

Full list of author information is available at the end of the article



They recognize virally encoded fake MHC self-molecules that are expressed by virus infected cells thus overcoming viral immune escape by NK cell activation [1].

Other paired receptors directly interact with viral or bacterial pathogens [1]. Among those are SIRP $\alpha$  and CEACAM1 and CEACAM3, members of the human carcinoembryonic antigen-related cell-cell adhesion molecule (CEACAM) family which have inhibitory ITIM/ITSM motifs and activating ITAM-like motifs in their cytoplasmic regions, respectively [3, 4]. A number of bacterial pathogens like pathogenic *Neisseria* (*N. gonorrhoeae*, *N. meningitidis*) *Haemophilus influenzae* and *Moraxella catarrhalis* have been shown to bind to the N-terminal immunoglobulin (Ig) variable-like domain of CEACAM1 on epithelial and immune cells allowing both entry into the host by transcytosis and down-regulation of the host's immune response by inhibiting adaptive and innate immune reactions [5–11]. Pathogens thus exploit the normal physiological function of CEACAM1 which acts as an immune inhibitory receptor on leukocytes upon homotypic or heterotypic interactions for example with other CEACAM members [7, 12]. In contrast, binding to granulocyte-specific CEACAM3 leads to uptake and destruction of these pathogens by triggering bactericidal processes [13–16]. Interestingly, phylogenetically unrelated adhesins such as opacity-associated (Opa) protein, outer membrane protein P5 and ubiquitous surface protein (UspA1) mediate interaction with the pathogen receptor CEACAM1 indicating convergent evolution [17–19].

A host-pathogen arms race involving receptors and decoy receptors with very similar adhesin-binding domains should lead to selection of pathogens with preferential binding to the inhibitory receptor and reduced binding to its decoy counterpart. Indeed, clinical isolates of *N. gonorrhoeae* from male urethra and female genital tract often express Opa proteins which bind to CEACAM1 but not to CEACAM3 [20]. The capability of *Neisseria* to randomly switch on expression of variant Opas from a panel of Opa genes aides natural selection from a heterogeneous *Neisseria* population. On the other hand, individuals with variant CEACAM1 receptors with low or no binding to pathogens should have an selective advantage. This will inevitably lead to poorly matched paired receptors and loss of decoy function. Intrachromosomal recombination or gene conversion between exons encoding ligand-binding domains of inhibitory and activation receptors within the CEACAM gene cluster could correct this deficit. Indeed, replacement of part of CEACAM3 exon 2 encoding the ligand-binding domain with sequences from the corresponding exon of CEACAM1 has happened in humans [3, 21].

CEACAM families differ greatly in gene number and domain composition of the encoded proteins between

mammalian species. Most of the analyzed mammals also contain putative paired CEACAM receptors [3, 22]. Allelic variants of CEACAM1 in mice and cattle have been shown or are suspected to serve as coronavirus receptors [23, 24]. Therefore, the rapid divergence of CEACAM1 and corresponding activating receptors during mammalian evolution is thought to be pathogen-driven [1, 3, 22].

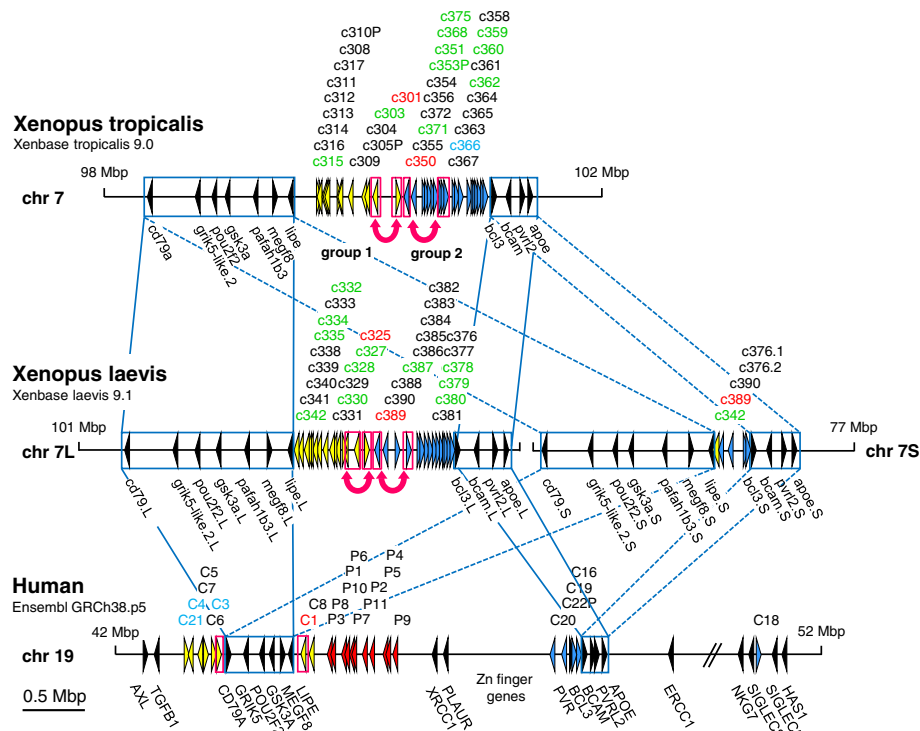
Also more distantly related CEACAM genes exist in mammals (CEACAM16, CEACAM18, CEACAM19 and CEACAM20) which do not represent paired receptors. They are clustered distally from the CEACAM1-related genes. They differ in domain organization and sequence among each other and exhibit specialized functions [25–27]. However, they are conserved between mammalian species which allows unequivocal assignment of orthologs [3].

CEACAM gene families seem to be restricted to vertebrates. CEACAM family members have been recently identified in reptiles, amphibians and in bony and cartilaginous fishes [28, 29]. However, the exact composition, the presence of paired receptors and the driving forces behind their evolution have not been investigated. Here we present comprehensive analyses of the *ceacam* families of two clawed frog species; the western clawed frog *Xenopus tropicalis* and the African clawed frog *X. laevis* the ancestors of which split some 60 million years ago [30]. We identified two distantly related *ceacam* families which both contain rapidly evolving paired receptors. Analysis of the *ceacam* family in *X. laevis* allowed us to follow the fate of a group of rapidly evolving genes after allotetraploidization.

## Results

### Identification of *ceacam* gene families in *X. tropicalis* and *X. laevis*

Based on their syntenic location between the flanking genes *lipo* and *bcl3*, and the presence of exons with conserved phasing encoding Ig variable (IgV)- and Ig constant (IgC)-like domains and ITIM and ITAM-like motifs most similar to mammalian CEACAM members (Fig. 1 and Additional file 1) 44 and 38 *ceacam* genes were identified on chromosomes 7 in *X. tropicalis* and *X. laevis*, respectively (Additional files 2 and 3). Interestingly, two *ceacam* gene loci exist in *X. laevis* on the homeologous chromosomes 7 L and 7S generated during speciation by hybridization of closely related species (Fig. 1; for nomenclature see [31]). Amino acid sequence comparison of the N-terminal IgV-like domains (N domains) revealed the presence of two distantly related subgroups group 1 and group 2 in both species (Fig. 2). N domains were chosen because they represent functionally important domains which have been shown in other species to be responsible for ligand binding [32]. Group 1 and group 2 genes are localized in clusters next to each other and, different from mammals, are not disrupted by *non-ceacam* genes (Fig. 1).



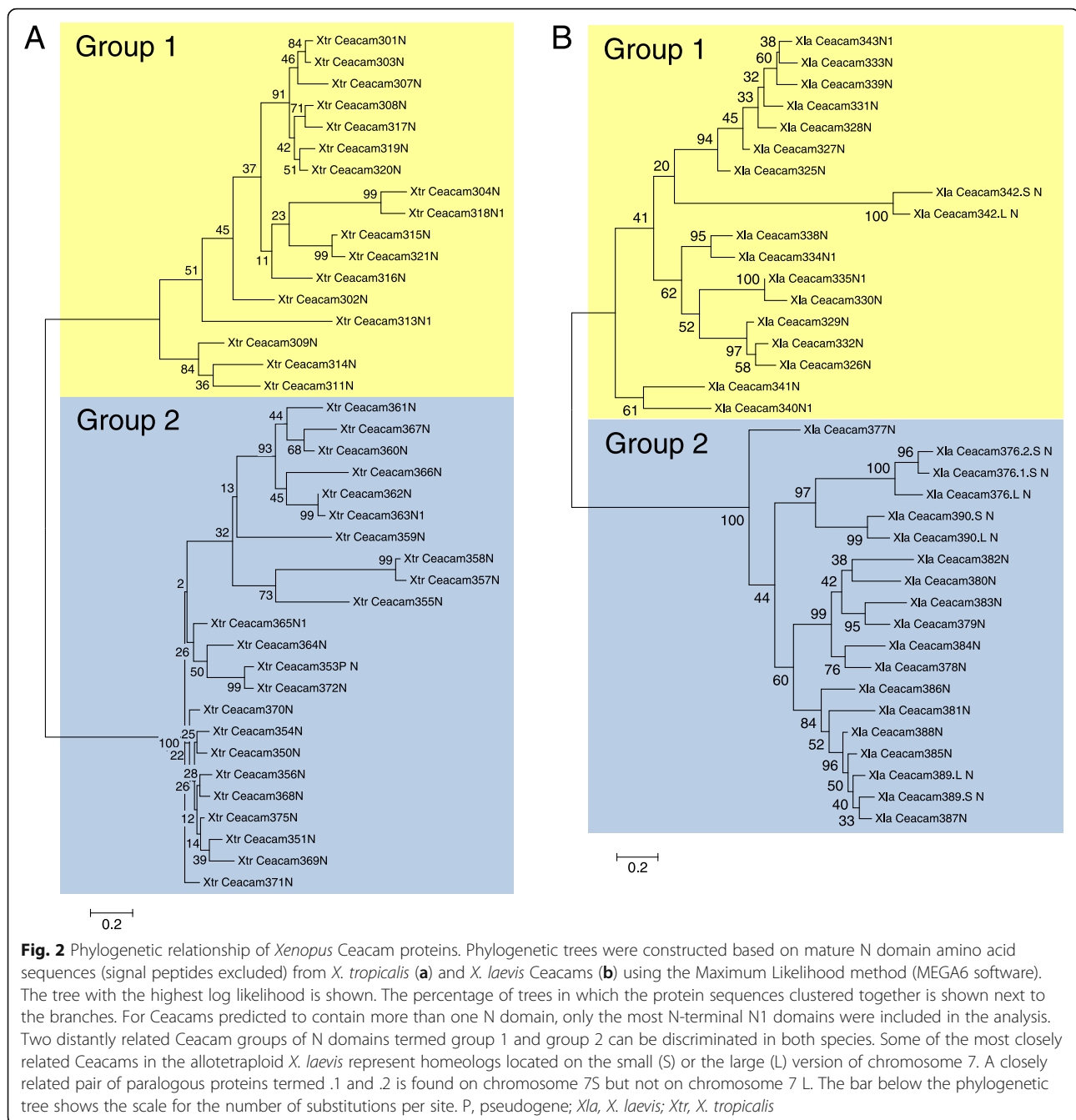
**Fig. 1** Chromosomal arrangement of *ceacam* genes in *X. tropicalis* and *X. laevis*. Arrowheads represent genes with their transcriptional orientation. *Xenopus ceacam* gene clusters are subdivided in group 1 (yellow) and group 2 members (blue). For *X. laevis* two homeologous *ceacam* clusters are found on chromosome 7 L and 7S which were generated by hybridization during speciation. Homeologs are indicated by L and S. Human *CEACAM1*-related genes are indicated in yellow, when predominantly expressed in trophoblast cells in red; the *CEACAM* genes for which orthologs can be identified in mammals are shown in blue and marker genes in black. Their syntenic relationship is shown by blue lines. Names of *CEACAM1*-like genes with ITIM-encoding exons are shown in red and with ITAM and ITAM-like motif-encoding exons in green and blue, respectively. Red double arrows symbolize potential recombination events between regions with *ceacam* genes containing ITIM or ITAM/ITAM-like motifs (genes shown in red boxes). Note that, in general, the genes in the same subcluster show the same, in different subclusters opposite transcriptional orientation. The nucleotide numbering of the chromosomes starts at the telomere of the short arms located to the left. The databases and the versions used are indicated below the species name. c, *ceacam*; C, *CEACAM*; chr, chromosome; P, pregnancy-specific glycoprotein (PSG) genes

Group 1 and group 2 Ceacam N domain amino acid sequences are most closely related to cartilaginous and bony fish and reptile and mammalian CEACAM N domain sequences, respectively (~35 % identity). Within subgroup 1 and 2, members exhibit between 40 and 93 % N exon amino acid sequence identity, while between subgroups only 20-30 % sequence identity is observed (data not shown). Similarly, transmembrane and cytoplasmic sequences also exhibit a higher degree of identity within groups than between groups (Additional file 1 and data not shown). Despite the low sequence identity group 1 and group 2 Ceacam IgV-like domains exhibit a very similar three-dimensional structure predicted by modeling using corresponding human and murine CEACAM1 sequences (Additional file 4).

Taken together, this indicates that two *ceacam* groups exist in *Xenopus* which were probably derived early in amphibian evolution possibly from two different *ceacam* ancestors and their origin predates the divergence of *X. laevis* and *X. tropicalis*.

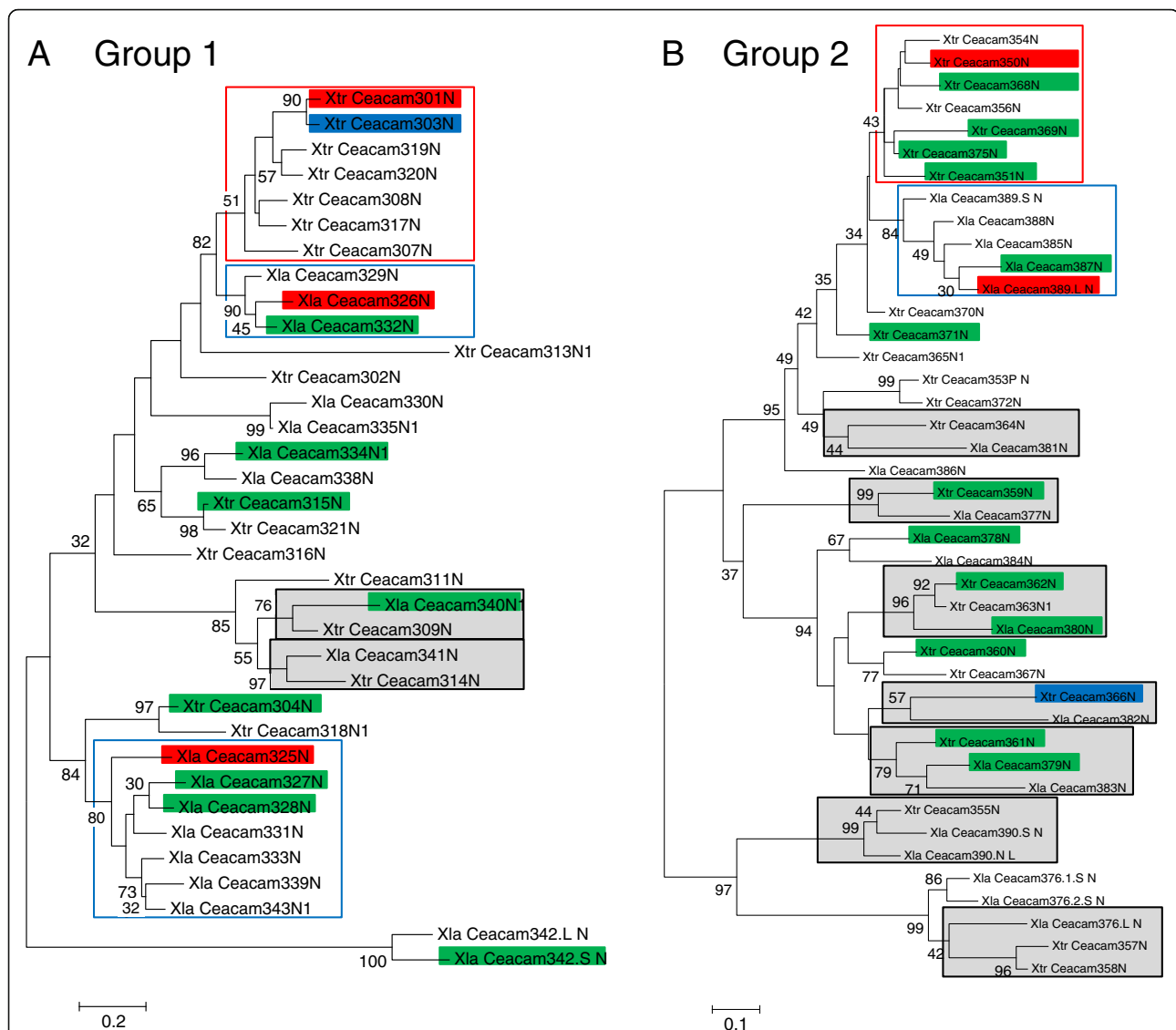
### Groups of paralogous Ceacams contain paired receptors

In mammals, CEACAM families consist of a group of orthologous members (CEACAM16, CEACAM18, CEACAM19, CEACAM20) where counterparts can be clearly assigned in different species and a group of paralogous members which are most closely related to CEACAM1 within the same species [3]. To identify orthologous Ceacam pairs as well as Ceacam paralogs, *X. tropicalis* and *X. laevis* group 1 and group 2 amino acid sequences from mature N domains (signal peptide sequence removed) were compared and their relationship displayed as dendrograms. In group 1 and group 2, two and seven pairs of orthologous Ceacams, respectively, could be identified based on their degree of sequence identity (Fig. 3). Their predicted domain organization is heterogeneous. Three members consist of only one IgV-like domain and are either secreted or membrane-bound by a transmembrane domain or a GPI anchor while two transmembrane-bound orthologous pairs are composed of one IgV- and one IgC-like domain and an ITAM-containing cytoplasmic region (Fig. 4).



In addition, sets of proteins exist whose closest relatives are found in the same species thus representing paralogous proteins: one and two in *X. tropicalis* and *X. laevis* group 1 Ceacams, respectively, and one in each species in group 2 (Fig. 3). Interestingly, these groups of closely related Ceacam members harbor one member with an ITIM and one or more with ITAM-like motifs (Fig. 3; Additional file 1). Pairs of cell surface proteins with similar extracellular domains which are able to interact with the same ligand, however, transmitting opposing i.e. inhibitory or activating signals represent so

called paired receptors. Based on these definition, in *X. tropicalis* Ceacam301 and Ceacam350 with an ITIM and Ceacam303 and Ceacam351 or Ceacam368 with an ITAM-like motif and an ITAM, respectively, and correspondingly in *X. laevis* Ceacam325 and Ceacam326 (ITIM) and Ceacam327 or Ceacam328 and Ceacam332 (ITAM) and Ceacam389 (ITIM) and Ceacam387 (ITAM) could represent paired receptors (Figs. 3 and 4). These putative paired receptors share between 80 and 93 % of their IgV-like ligand binding domain amino acid sequences (data not shown).



**Fig. 3** Orthologous and paralogous members of the Ceacam family in *X. tropicalis* and *X. laevis*. Phylogenetic trees were constructed based on mature N domain amino acid sequences (signal peptides excluded) from *X. tropicalis* and *X. laevis* group 1 (**a**) and group 2 Ceacams (**b**) using the Maximum Likelihood method (MEGA6 software). The tree with the highest log likelihood is shown. The percentage of trees in which the protein sequences clustered together is shown next to the branches. For Ceacams predicted to contain more than one N domain, only the most N-terminal N1 domains were included in the analysis. Ceacams whose N domains exhibit the highest degree of identity within the same species (paralogs) are boxed with red and blue lines for *X. tropicalis* and *X. laevis*, respectively. Pairs consisting of most closely related proteins from different species (sometimes including recently generated paralogs) are highlighted in gray (orthologs). The names of ITIM-containing members are highlighted in red, proteins with ITAM and ITAM-like motifs are marked with green and blue background, respectively. P, pseudogene; Xla, *X. laevis*; Xtr, *X. tropicalis*

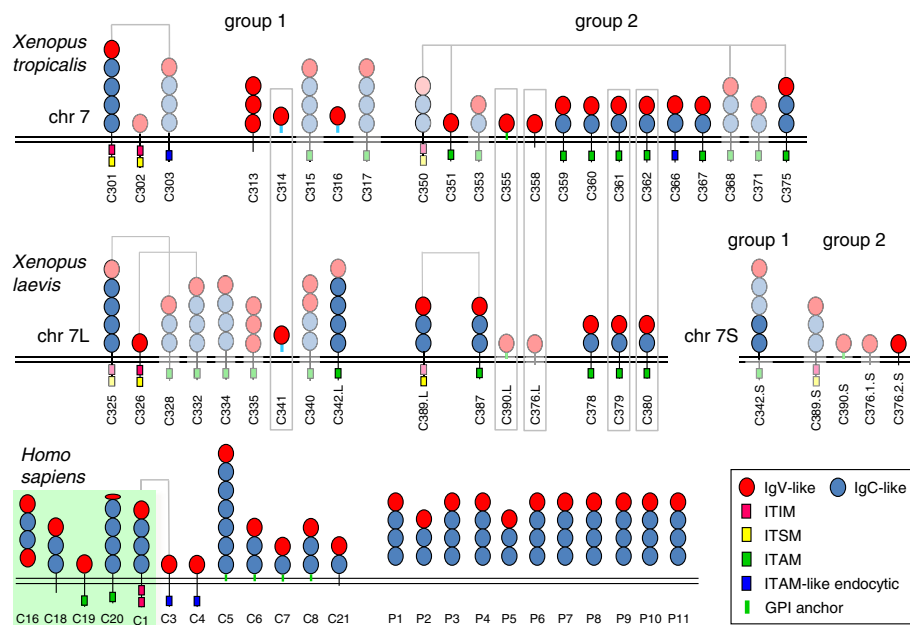
In summary, orthologous and paralogous members exist in both Ceacam groups. Putative paired receptors could be identified among the paralogous members.

#### Similarity of receptor binding domains of paired receptors is maintained by recombination

In paired receptors, pathogen-binding regions have to stay similar thus allowing the host to counterbalance the immune suppressive signal elicited by engagement of the inhibitory receptor through the pathogen by providing an

activating receptor as a mimic [33, 34]. In other paired receptor gene families this is often achieved by recombination or gene conversion between genes encoding inhibitory and activating receptors, restricted to the gene region encoding the ligand binding domains [33]. We, therefore, screened the potential *Xenopus* paired receptor genes for recombination/gene conversion events. Indeed, in *X. tropicalis* group 1 *ceacam301* and *ceacam303* and in group 2 *ceacam350* and *ceacam351/ceacam368* N exons have recently undergone gene conversion which is exactly





**Fig. 4** Domain organization of *Xenopus* Ceacam proteins. The domain organization of Ceacam family members from *X. tropicalis* and *X. laevis* was predicted by gene analysis only (domains shown in *light colors*) or, where available, confirmed by EST and cDNA sequences (domains shown in *intense colors*). Only Ceacams whose complete domain organization could be delineated were included. Homeologous Ceacams in *X. laevis* arisen by hybridization during speciation are discriminated by L (encoded on long chromosome variant 7) and S (encoded on short chromosome variant 7). For comparison the domain organization of human CEACAM proteins is shown. The conserved human CEACAM family members are highlighted in *green*. IgV-like domains are shown as *red*, IgC-like domains are *blue* ovals. The predicted signaling motifs in the cytoplasmic domains are schematically shown as *green* (ITAM), *blue* (ITAM-like motif), *red* (ITIM) and *yellow boxes* (ITSM). Transmembrane domains and GPI anchors are indicated by *black* and *green lines*, respectively. Orthologous relationship as suggested by sequence relatedness and/or synteny is indicated by *gray boxes*. Note that Xtr\_Ceacam301 and Xla\_Ceacam326 appear to represent orthologs based on the degree of N domain sequence identity (see Fig. 3) despite their different domain organization. Paired receptors identified by their similar N domain sequences and the presence of ITIM/ITSM or ITAM/ITAM-like motifs are connected by *gray lines*. C, CEACAM or Ceacam; P, PSG

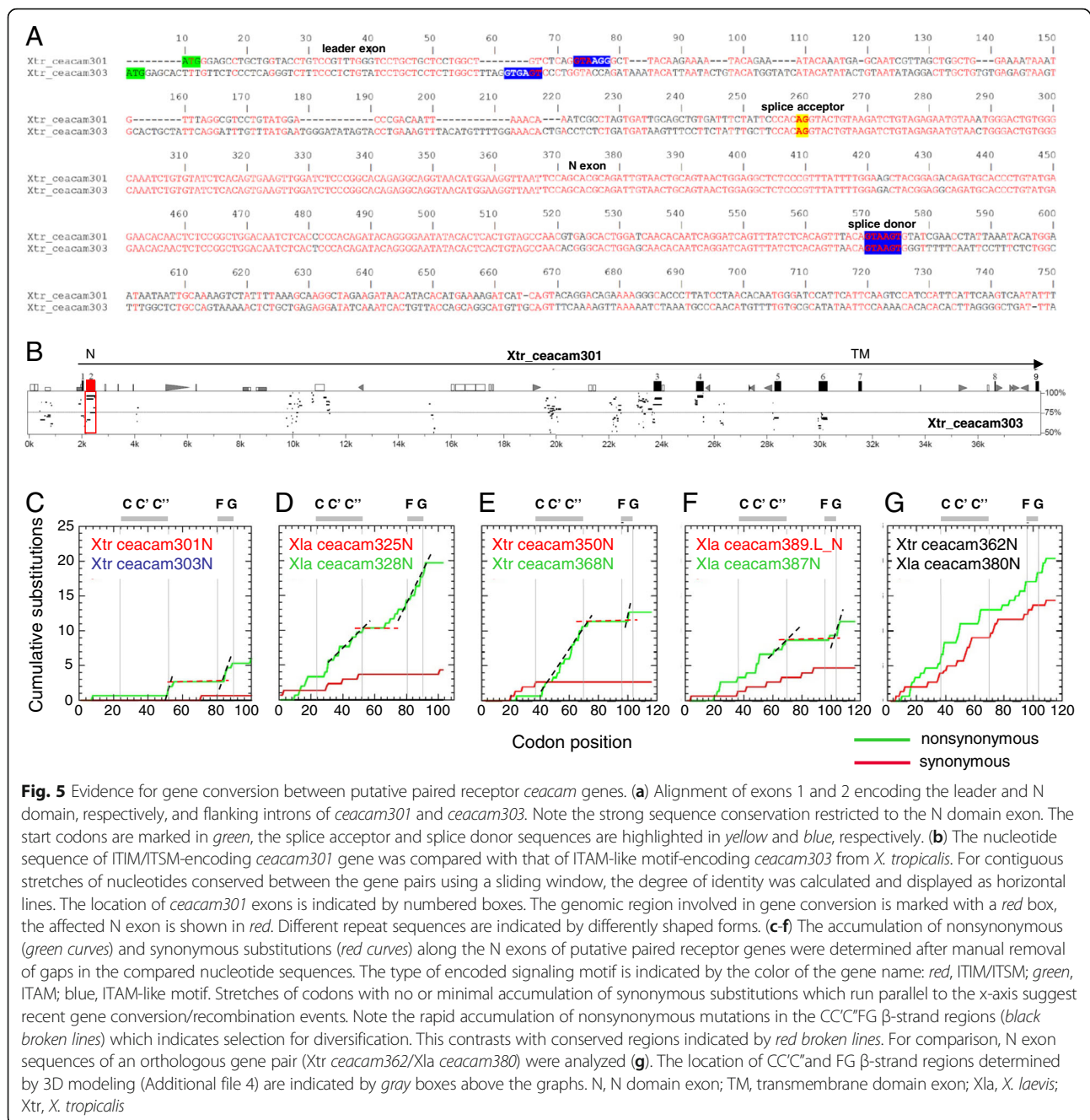
restricted to the exon just including the splice consensus sequences. This is supported by the high conservation of N exon sequences with virtual absence of synonymous mutations (which in general do not encounter purifying selection) and very low sequence conservation in the flanking introns (Fig. 5a, b, c; Additional file 5). No other genomic regions seem to have been involved in the gene conversion event in *ceacam301* and *ceacam303* (Fig. 5b). Recombination events were also noticed in N domain exons of other putative paired receptor gene pairs. This was evident from the lack or low rate accumulation of synonymous mutations mostly restricted to certain regions of the N exons of receptor pairs like *ceacam325* and *ceacam328* and possibly of *ceacam389* and *ceacam387* in *X. laevis* and *ceacam350* and *ceacam368* in *X. tropicalis* (Fig. 5d-f). Without recurrent recombination/gene conversion events between paralogous genes one would expect a steady accumulation of synonymous nucleotide changes along the N exons as it is found between *X. tropicalis* and *X. laevis* *ceacam* orthologs, like *ceacam362* and *ceacam380* (Fig. 5g). Interestingly, possibly due to continued pressure from pathogen adhesion-receptor interaction, regions which represent putative interaction sites (the CC'CFG face of the Ig fold [35]) appear to

rapidly accumulate non-synonymous mutations occurring after a gene conversion event (Fig. 5c-f). For intrachromosomal recombination/gene conversion to take place, involved genes must exhibit opposite transcriptional orientation in order that homologous sequences can be aligned with looping out of the intervening sequences [3, 36]. Indeed, putative paired receptor genes in *X. laevis* and *X. tropicalis* from both *ceacam* subgroups exhibit opposite transcriptional orientation with group 1 and group 2 ITIM-encoding genes facing each other (Fig. 1).

The non-random transcriptional orientation of *ceacam* genes encoding proteins with inhibitory or activating signaling motifs, gene conversion within exons encoding ligand binding domains as well as the conservation of these domains in ITIM and ITAM-containing Ceacams strongly argue that these Ceacams function as paired receptors.

#### Selection for diversification in paired receptor Ceacam groups

Pathogen receptors which allow entry into a host often exhibit selection for diversification of their amino acid sequences. This is evident from high ratios (>1) of their



rate of non-synonymous over their rate of synonymous mutations (dN/dS) especially in regions relevant for pathogen binding [37]. When whole domains or proteins are analyzed the dN/dS ratios can drop below 1 despite the presence of regions with strong positive selection, due to the presence of regions with negative or neutral selection. To test whether ITIM/ITSM-bearing Ceacam receptors in *Xenopus* might represent pathogen receptors as found for human and mouse CEACAMs we analyzed dN/dS ratios of N domain exons of *ceacam* orthologs in *X. tropicalis* and *X. laevis*. Orthologous

genes encoding receptors with ITIM exhibited the highest dN/dS ratios, i.e. 1.3 for *Xtr ceacam301*/*Xla ceacam326* (group 1) and 1.0 for *Xtr ceacam350*/*Xla ceacam389.L* orthologs (group 2) (Fig. 3; Fig. 6a). In contrast, dN/dS ratios between 0.18 and 0.65 were found for the other *ceacam* orthologs. The lowest dN/dS ratios were observed for the flanking non-*ceacam* genes (dN/dS = 0.1–0.25) with the exception of the immune function gene *cd79a* (dN/dS = 0.35), which encodes an ITAM-bearing component of the B cell receptor. The orthologous *ceacam* pair with the lowest dN/dS ratio (dN/dS = 0.18)

encodes glycosylphosphatidylinositol (GPI) membrane-anchored proteins (Fig. 6a). No large dN/dS ratio differences were found when *X. tropicalis* genes were compared with the *X. laevis* homeologs on chromosome 7S (Fig. 6a) which are paralogous genes that were formed by the hybridization event in *X. laevis* which probably occurred during speciation (see below). This suggests that there is no loss of function or gain of new function for one of the two homeologs.

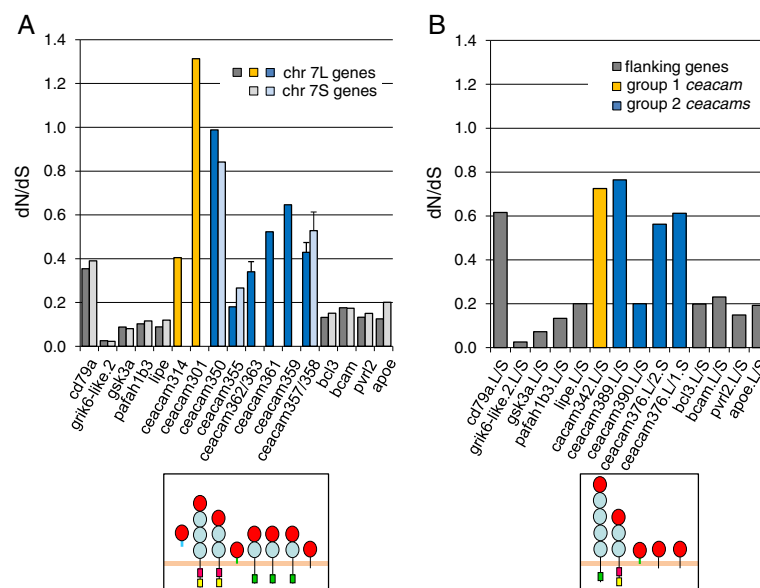
N exon dN/dS ratios of  $\geq 1$  of *Xenopus* inhibitory receptor orthologs indicate selections for diversification typically observed in pathogen receptors.

#### Selective gene loss in *ceacam* locus on chromosome 7S created by allotetraploidization in *X. laevis*

In *X. laevis*, allotetraploidy was probably caused by hybridization during speciation which led to whole genome duplication including the *ceacam* gene locus on chromosome 7. Despite the overall conservation of the duplicated genomic region surrounding the *ceacam* loci only the locus on chromosome 7 L resembles that found in *X. tropicalis*. More than 80 % of *ceacam* genes were lost from the locus on chromosome 7S and only 1 and 3 genes are left from group 1 and group 2 genes, respectively (Fig. 1). Interestingly, no paired receptor pair is preserved at the 7S locus; only the group 2 inhibitory receptor gene

(*ceacam389.S*) is retained. Notwithstanding the massive *ceacam* gene loss, a gene duplication event probably after *X. laevis* speciation led to the generation of two closely related *ceacam376* genes (*ceacam376.1.S* and *ceacam376.2.S*) next to the *bcl3* gene (Fig. 1).

To determine whether selection occurs to maintain the function of both homeologous gene copies, dN/dS ratios for N domain exons or whole coding sequences were determined for *ceacam* and flanking genes, respectively. Only one *ceacam* gene which encodes a GPI-linked group 2 member reveals similar conservation with a dN/dS ratio of 0.2 as flanking genes which exhibit dN/dS ratios between 0.03 and 0.2 with the exception of *cd79a* exhibiting a dN/dS ratio of 0.6 (Fig. 6b). In addition, both copies of all homeologous genes exhibit similar dN/dS ratios when compared with the *X. tropicalis* orthologs (Fig. 6a). Therefore, both homeologous copies of *ceacam390.L/390.S* and the non-*ceacam* genes seem to be functional and under selective pressure. In contrast, much higher dN/dS ratios between 0.6 and 0.75 are observed for the remaining *ceacam* genes, indicating either lack of selection for conservation of the homeologous pairs or selection for diversity. The latter appears to be at work for the inhibitory receptor-encoding genes *ceacam389.L/389.S* which exhibit the highest dN/dS ratios (Fig. 6a, b). In addition, regions with increased accumulation of non-synonymous mutations



**Fig. 6** Differential conservation of orthologous and homeologous *Xenopus ceacam* and flanking genes. Nucleotide sequences of N exons and the open reading frames from *ceacam* and flanking genes, respectively, from *X. tropicalis* were compared codon-wise with each of the two homeologous orthologs of *X. laevis* (a) or the *X. laevis* homeologs with each other (b) after manual removal of gaps and the ratio of the rate of nonsynonymous and synonymous mutations was calculated. The resulting dN/dS values were plotted using the *X. tropicalis* gene order on chromosome 7 (a) or the *X. laevis* gene order on chromosome 7S (b). In cases of recent gene duplication of one of the orthologs in *X. tropicalis* or *X. laevis* dN/dS ratios for all orthologous pairs were calculated and plotted as mean and deviation or standard deviation. The type of signaling motif present in the corresponding proteins can be inferred from the schematic representation of the protein domains below the graphs (see Fig. 4 for keys to domains and motifs). Chr, chromosome



which are similar to those found for orthologous or homeologous pairs of inhibitory receptors could be identified. They co-localize with the CC'C''FG  $\beta$ -sheet which represent putative pathogen binding sites (Fig. 7).

## Discussion

### Amphibian *ceacam* gene family exhibits an ancestral genomic arrangement

In this comprehensive analysis *ceacam* genes could be identified in *X. tropicalis* and *X. laevis* based on synteny and structural homology despite a low degree of sequence identity which is due to rapid divergence during evolution. The *ceacam* gene family consists of two distantly related subgroups (group 1 and group 2) with 15–20 members each. Only few orthologous *ceacam* members (mostly in group 2) were found in the two frog species. The *Xenopus ceacam* genes are arranged in one cluster separated by subgroups uninterrupted by non-*ceacam* genes (Fig. 1). This is also observed for most of the *CEACAM* genes in the marsupial opossum (*Monodelphis domestica*) but not in eutherian mammals. Here, the *CEACAM* locus is interrupted by two large regions with non-*CEACAM* genes (Fig. 1; [3]). This indicates that a continuous *ceacam* gene cluster was also present in the last common ancestor of amphibians and mammals.

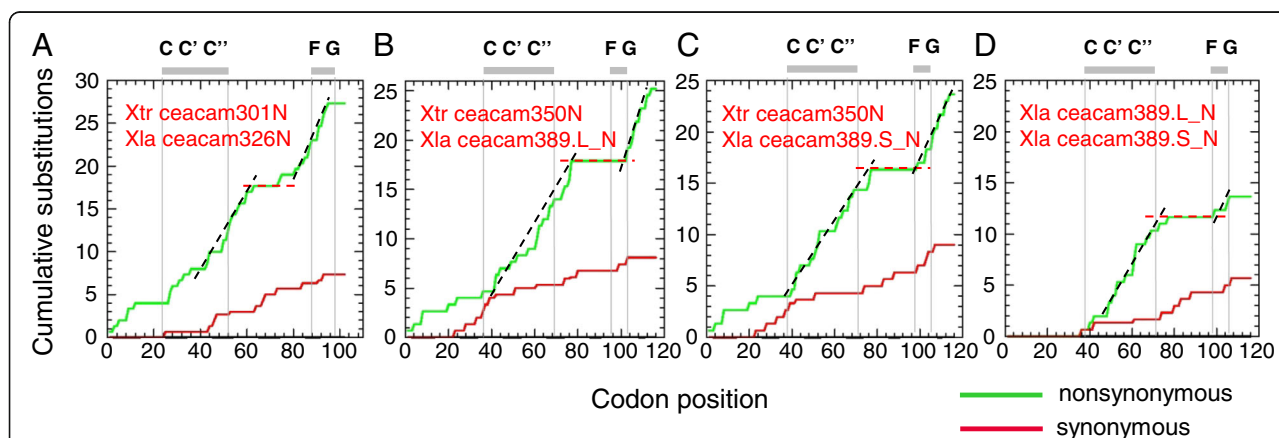
No *CEACAM16*, *CEACAM18*, *CEACAM19* and *CEACAM20* genes, which are well conserved in mammals and are clustered next to *BCL3*, were found in *Xenopus* [3]. However, *CEACAM19* orthologs can be identified unequivocally in reptiles including turtles, snakes, alligators and gecko but not the other genes ([29]; Zimmermann, unpublished results). Interestingly, group 2 *Xenopus*

*ceacam* genes which are also located next to *bcl3* but not group 1 genes are most closely related to reptilian *CEACAM19* (32–37 % identity) representing the most common hits outside of the anuran order when *Xenopus ceacam* group 2 N exon amino acid sequences are used as query sequences. This might indicate that *CEACAM19* in reptiles and mammals and group 2 *ceacam* genes share a common ancestor.

### Two paired receptor systems exist in the *Xenopus ceacam* family which exhibit signs of pathogen-mediated selection

Each subgroup contains one or two paired receptors with oppositely signaling ITIM and ITAM/ITAM-like motifs and highly similar ligand binding domain (N domain) amino acid sequences (Figs. 3 and 4). This is different from mammalian *CEACAMs* which typically have only one set of paired receptors or none as found for mouse and rat [3].

What is the evidence that the paired *Ceacam* receptors are being or have been used as pathogen receptors in *Xenopus*? Both diversification of pathogen receptors to avoid binding of the pathogen (indicated by high dN/dS ratios) and maintenance of similarity of the pathogen adhesin-interacting domain in paired receptors functioning as pathogen or decoy receptors will be selected for in a pathogen/host arms race [1, 34, 37]. Indeed this is observed in both *Xenopus ceacam* groups. The exons encoding the ligand-binding domains of *X. tropicalis* and *X. laevis* ITIM-containing orthologs exhibit the highest dN/dS ratios which is indicative of positive selection (Fig. 6a). Positively selected amino acid positions of a



**Fig. 7** Selection for sequence diversification in inhibitory receptor *Ceacams*. The accumulation of non-synonymous (green curves) and synonymous substitutions (red curves) along the N exons of orthologous (a–c) and homeologous inhibitory receptor gene pairs (d) were determined after manual removal of gaps in the compared nucleotide sequences. The names of the analyzed genes are indicated in the top left corner of the graph. Note the rapid accumulation of nonsynonymous mutations in the CC'C''FG  $\beta$ -strand regions (black broken lines) which indicates selection for diversification. This contrasts with conserved regions between CC'C'' and FG  $\beta$ -strands indicated by red broken lines. The location of CC'C'' and FG  $\beta$ -strand regions determined by 3D modeling (Additional file 4) are indicated by gray boxes above the graphs. Note the sequence identity of the 38 N terminal codons in the N exon of *X. laevis* homeologs *ceacam389.L* and *ceacam389.S* possibly caused by interchromosomal recombination. N, N domain exon; Xla, *X. laevis*; Xtr, *X. tropicalis*

protein domain are expected to reside at the site of contact between the pathogen adhesin and its receptor [37]. This seems to be the case for putative Ceacam pathogen receptor orthologs in *X. tropicalis* and *X. laevis* which show selective accumulation of non-synonymous mutations in the CC'C" FG face of the IgV-like domain (Fig. 7a-c) responsible for pathogen interaction [35]. In contrast, the Ig  $\beta$ -sheet on the opposite side of the Ig fold is highly conserved. This recurrent host escape and pathogen adaptation ("Red Queen" scenario [38]) can lead to an imbalance with pathogen binding to the entry/inhibitory receptor being maintained while the decoy receptor-pathogen interaction is abolished [20]. This problem can be resolved through recombination/gene conversion whereby the decoy receptor pathogen binding domain is replaced by that of the pathogen receptor. Absence of synonymous mutations in large regions of the N exons revealed by comparison of the putative pathogen receptor and decoy receptor sequences suggest intra- or interchromosomal recombination or gene conversion events as cause of the similarity of paired receptor ligand binding domains (Fig. 5c-e). Original sequence identity is rapidly lost due to the on-going struggle between host and pathogens again noted by the selective accumulation of non-synonymous mutations in the presumed adhesin binding regions (Fig. 5c-e).

Comparative analyses of mammalian *CEACAM* loci revealed inversion of regions with non-*CEACAM* genes between oppositely oriented *CEACAM1* and presumed decoy receptors genes in some species. This suggested that recombination involved an intrachromosomal loop formation mechanism that allows alignment of the exon sequences encoding the ligand binding domains [3]. The recombination mechanism in *Xenopus* is still unclear. However inverted transcriptional orientation of ITIM- and ITAM-bearing genes indicates a similar intrachromosomal recombination mechanism.

#### Loss of paired receptors from one of the two homeologous *X. laevis* ceacam loci

At the time of hybridization of two ancestral *X. laevis* species some 40 million years ago [30] which led to speciation, two probably functionally distinguishable sets of *ceacam* genes existed. The set on chromosome 7 L was better functioning presumably with respect to pathogen resistance and was consequently retained [39]. Loss of the *ceacam* locus on chromosome 7S is not complete. Interestingly, in both subgroups no paired receptor system persisted. Only one ITIM- and one ITAM-bearing Ceacam is found in group 2 and group 1, respectively (Fig. 6b). This indicates that once the genes for inhibitory or activating members have been lost the remaining gene cluster degrades rapidly probably due to lack of

selection pressure exerted by pathogens which helps to maintain paired receptors.

#### Conclusions and Perspectives

Although we do not know which specific pathogens bind to Ceacam receptors in *Xenopus* the presence of closely related paired receptors as well as selection for diversification suggests that also in amphibians CEACAM1-related inhibitory proteins still are or have been exploited in the past as pathogen receptors and similar defense strategies have been developed in amphibians and mammals by convergent evolution. Thus the *CEACAM* family is a prototype gene family which offers a unique opportunity to study "arms races" caused by host/pathogen interactions.

CEACAM1-related pathogen receptors serve a dual role: They allow both entry into the host and inhibition of inflammatory responses to pathogen infections. Therefore, CEACAM1-like receptors are expected to be expressed on epithelial surfaces as well as on leukocytes involved in innate and adaptive immunity. Decoy receptors should be expressed on cells of the innate immune system which allow uptake and destruction of pathogens. The identification and characterization of individual members of the *Xenopus ceacam* gene families will now allow to use next generation sequencing data of RNA from multiple organs and cell types of *Xenopus* species as well as genomic sequence data of additional frog species like that of *Nanorana parkeri* a member of the species-rich *Neobatrachia*, which contains the vast majority of amphibian taxa (Sun et al., 2015) to support the suggested pathogen defence function of anuran *ceacam* families.

#### Methods

##### Identification and nomenclature of genes

Sequence similarity searches were performed using the NCBI BLAST tools (<http://www.ncbi.nlm.nih.gov/BLAST>) and the Ensembl BLAST/BLAT ([http://www.ensembl.org/Xenopus\\_tropicalis/Tools/Blast?db=core](http://www.ensembl.org/Xenopus_tropicalis/Tools/Blast?db=core)) and Xenbase BLAST (<http://www.xenbase.org/genomes/blat.do>) search programs. For identification of *ceacam* genes regions syntenic to mammalian *CEACAM* loci were analyzed for the presence of Ig domain-encoding genes. The following databases were used for *ceacam* gene identification and loci analyses: Xenbase *X. laevis* J-Strain 9.1 and *X. tropicalis* Nigerian 9.0 (<http://www.xenbase.org/entry/>) and Ensemble *X. tropicalis* JGI 4.2 ([http://www.ensembl.org/Xenopus\\_tropicalis/Info/Index](http://www.ensembl.org/Xenopus_tropicalis/Info/Index)). *Xenopus* genomes were reprobbed with exon sequences from newly discovered *ceacam* genes. For estimation of the number of *ceacam* genes present in a given species, distinct *ceacam* N domain exons with a sequence divergence > 1 % were counted. Multiple N exons with no annotated non-N exon in between were considered

to belong to the same gene. Genes that contained stop codons within their N domain exons or lacked appropriate splice acceptor and donor sites were considered to represent pseudogenes. Genes were assigned to their respective *ceacam* subgroups 1 and 2 based on phylogenetic analyses. Due to their non-orthologous relationship with mammalian CEACAM genes the new *ceacam* genes were numbered independently as follows: *X. tropicalis*, group 1: *ceacam301-ceacam321*; *X. tropicalis*, group 2: *ceacam350-ceacam375*; *X. laevis*, group 1: *ceacam325-ceacam342*; *X. laevis*, group 2: *ceacam376-ceacam390*. Nucleotide sequences from the N domain exons can be used as gene identifier (Additional files 2 and 3). Gene Nomenclature Guidelines recommended by Xenopus Gene Nomenclature Committee (2013) was followed (<http://www.xenbase.org/gene/static/geneNomenclature.jsp>).

### Sequence motif identification and 3D modeling

The presence of ITAM, ITAM-like and ITIM/ITSM motifs were confirmed using the amino acid sequence pattern search program ELM (<http://elm.eu.org/>). Transmembrane regions, glycosylphosphatidylinositol (GPI) signal domains and leader peptide sequences were identified using the TMHMM (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>), the big-PI predictor ([http://mendel.imp.ac.at/sat/gpi/gpi\\_server.html](http://mendel.imp.ac.at/sat/gpi/gpi_server.html)), GPI-SOM (<http://gpi.unibe.ch/>) and the SignalP 4.1 programs (<http://www.cbs.dtu.dk/services/SignalP/>), respectively [40]. For three-dimensional modeling the geno3D software was used ([https://geno3d-prabi.ibcp.fr/cgi-bin/geno3d\\_automat.pl?page=/GENOHLP/genohlp\\_help2.html](https://geno3d-prabi.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENOHLP/genohlp_help2.html)). Images were constructed with the Swiss-PdbViewer software 4.1.

### Phylogenetic analyses and determination positive selection

Phylogenetic analyses based on amino acid sequences were performed with the MEGA6 package [41]. The applied Maximum Likelihood method is based on the JTT matrix-based model [42]. The trees with the highest log likelihood are depicted. The percentage of trees in which the protein sequences clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. All positions containing gaps and missing data were eliminated. In order to determine the selective pressure on the maintenance of the nucleotide sequences, the number of nonsynonymous nucleotide substitution per nonsynonymous site (dN) and the number of synonymous substitutions per synonymous site (dS) were determined for N domain exons. The dN/dS ratios as well as the cumulative synonymous and nonsynonymous substitutions along coding regions of N domain exons from paralogous and

orthologous genes were calculated after manual editing of sequence gaps or insertions guided by the amino acid sequences using the SNAP program (Synonymous Nonsynonymous Analysis Program; <http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html>). The program *Pip-Maker* (<http://bio.cse.psu.edu/>) was used to identify conserved contiguous stretches of nucleotides between gene pairs and to calculate the degree of identity which is summarized as a 'percent identity plot' [43]. Multiple amino acid and nucleotide sequence alignments were performed with ClustalW programs ([http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_clustalw.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_clustalw.html)).

### Additional files

**Additional file 1:** ITIM and ITAM/ITAM-like sequence motifs in the cytoplasmic domains of *Xenopus* Ceacam proteins. The amino acid sequences encoded by cytoplasmic domain exons of putative inhibitory (A) or activating *X. tropicalis* and *X. laevis* Ceacams (B) were aligned. Intra- and inter-group 1 and group 2 Ceacam member alignments are shown. The amino acids in one letter code are colored according to their relatedness: red, identical; green, highly similar properties; blue, similar properties. Gaps indicate exon borders, dashes missing amino acids. The names of the cytoplasmic domain-encoding exons and the intron types (0, xxx-intron-xxx; 1, x-intron-xx; 2, xx-intron-x; xxx = codon) are indicated above and below the aligned sequences. ITIM (I/L/V/SxYxxL/V), ITSM (TxYxxV/I), ITAM (D/ExxYxxL/Ix<sub>6-8</sub>YxxL/I) and endocytic or ITAM-like motifs (YxxL/MV/I/F) are marked by red, orange, green and blue boxes, respectively. Note the characteristic split of the YxxL motif in the ITAM and ITAM-like motifs by phase 0 introns. GRB2-like Src homology 2 (SH2) domains-binding motifs (YxNx) are boxed with magenta lines. Note their conservation in most ITAM/ITAM-like motif-containing Ceacams at presumably non-homologous positions in group 1 and group 2. The small letter x represents any amino acid, and slashes separate alternative amino acids that may occupy a given position. An additional motif (TEHKS) highly conserved in mammals and shown to bind beta-catenin is shaded gray [44]. Note the presence of two types of dissimilar cytoplasmic sequences within (for ITAM-like motifs) and between Ceacam groups (ITIM and ITAM-like) differing in part also in the number and phasing of the encoding exons. Cyt, cytoplasmic domain exon; ITAM, immunoreceptor tyrosine-based activation motif; ITIM, immunoreceptor tyrosine-based inhibition motif; ITSM, immunoreceptor tyrosine-based switch motif; Mdo, *Monodelphis domestica* (gray short-tailed opossum); Xla, *X. laevis*; Xtr, *X. tropicalis*. (DOCX 95 kb)

**Additional file 2:** *Xenopus tropicalis ceacam* N exon nucleotide sequences. *Xenopus tropicalis ceacam* N exon nucleotide sequences (Xenbase tropicalis 9.0 Chr07 unless otherwise indicated, e.g. Ensembl *Xenopus* JGI 4.2). EST, expressed sequence tag; N, N or Ig variable-like domain exon; P, pseudogene (stop codon in N exon). (DOCX 23 kb)

**Additional file 3:** *Xenopus laevis ceacam* N exon nucleotide sequences. *Xenopus laevis ceacam* N exon nucleotide sequences (Xenbase laevis 9.1 unless otherwise indicated). EST, expressed sequence tag; N, N or Ig variable-like domain exon; P, pseudogene (stop codon in N exon). (DOCX 22 kb)

**Additional file 4:** Modeling of the three-dimensional structure of Ceacam301 and Ceacam350 IgV-like domains. Ceacam301 and Ceacam350 mature IgV-like domains were modeled using the geno3D software. The equivalent surface corresponding to the CC'C'FG face is shown as ribbons. The amino acids which flank the CC'C' and FG  $\beta$ -strands are indicated in three-letter code. Amino acids which belong to the CC'C'FG face of the sequence alignment are indicated in yellow. Human and murine CEACAM1 IgV-like domain sequences were used as template for modeling of Ceacam301 and Ceacam350 IgV-like domains, respectively. Note the overall similarity of the putative ligand-binding region shown at the left side despite the fact that the regions corresponding to the C'  $\beta$ -strand in CEACAM1 were not classified as  $\beta$ -strands in Ceacam301 and Ceacam350. (PPTX 190 kb)

**Additional file 5:** Sequence comparison of leader and N exon genomic regions from *Xenopus* paired receptor *ceacam* genes. Nucleotide sequences comprising N domain exons and flanking regions from presumed paired receptor *ceacam* genes (group 1 and group 2) as well as from the homeologous group 2 inhibitory receptor genes from *X. laevis* were pairwise aligned. Identical nucleotide positions are shown in red. Splice acceptor and donor consensus sequences are marked in yellow and blue, respectively, translational start codons in green. Note extended stretches of nucleotide sequences with high conservation restricted to N exon regions including splice sites but extending barely in to the introns. *Xla*, *X. laevis*; *Xtr*, *X. tropicalis*. (DOCX 83 kb)

## Abbreviations

ceacam: Carcinoembryonic antigen-related cell adhesion molecule; dN: Rate of non-synonymous substitutions; dS: Rate of synonymous substitutions; GPI: Glycosylphosphatidylinositol; Ig: Immunoglobulin; IgC: Ig constant; IgV: Ig variable; ITAM: Immunoreceptor tyrosine-based activation motif; ITIM: Immunoreceptor tyrosine-based inhibition motif; ITSM: Immunoreceptor tyrosine-based switch motif; MHC: Major histocompatibility antigens; NK: Natural killer cell; Opa: Opacity-associated; UspA1: Ubiquitous surface protein A1

## Acknowledgements

None.

## Funding

This work was supported by grants from GIZ (Contract no. 81170269), BMWI (KF2875802UL2) and DFG (HE 6249/4-1) to RK. No role in study design, in the collection, analysis, and interpretation of data, in the writing of the manuscript; and in the decision to submit the manuscript for publication was played by the funding bodies.

## Availability of data and materials

Nucleotide sequences from the N domain exons of the newly described *Xenopus ceacam* genes which can be used as gene identifiers for search in databases are provided (Additional files 2 and 3). Phylogenetic tree, sequence data and alignments used to produce the results displayed in Figs. 2 and 3 were deposited in TreeBASE (<https://treebase.org/treebase-web/home.html>). They are available using the following link <http://purl.org/phylo/treebase/phyloWS/study/TB2:S20145>.

## Authors' contributions

WZ conceived of the study, and participated in its design, carried out database searches and sequence alignments and drafted the manuscript. RK participated in the design of the study, performed the phylogenetic analyses and structural modeling. Both authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Consent for publication

Not applicable.

## Ethics approval

This study was exclusively computational. No *Xenopus* frogs or tissues were used.

## Author details

<sup>1</sup>Tumor Immunology Laboratory, LIFE Center, University Clinic, Ludwig-Maximilians-University, Feodor-Lynen-Str. 19, 81377 Munich, Germany. <sup>2</sup>Department of Urology, University Clinic, Ludwig-Maximilians-University, Marchioninistr. 15, 81377 Munich, Germany. <sup>3</sup>Institute of Immunology, Friedrich-Loeffler Institut, 17493 Greifswald-Insel Riems, Germany.

## References

- Akkaya M, Barclay AN. How do pathogens drive the evolution of paired receptors? *Eur J Immunol*. 2013;43(2):303–13.
- Martinet L, Smyth MJ. Balancing natural killer cell activation through paired receptors. *Nat Rev Immunol*. 2015;15(4):243–54.
- Kammerer R, Zimmermann W. Coevolution of activating and inhibitory receptors within mammalian carcinoembryonic antigen families. *BMC Biol*. 2010;8:12.
- Kuespert K, Pils S, Hauck CR. CEACAMs: their role in physiology and pathophysiology. *Curr Opin Cell Biol*. 2006;18(5):565–71.
- Boulton IC, Gray-Owen SD. Neisserial binding to CEACAM1 arrests the activation and proliferation of CD4+ T lymphocytes. *Nat Immunol*. 2002;3(3):229–36.
- Slevogt H, Zabel S, Opitz B, Hocke A, Eitel J, N'Guessan PD, Lucka L, Riesbeck K, Zimmermann W, Zweigner J, et al. CEACAM1 inhibits Toll-like receptor 2-triggered antibacterial responses of human pulmonary epithelial cells. *Nat Immunol*. 2008;9(11):1270–8.
- Gray-Owen SD, Blumberg RS. CEACAM1: contact-dependent control of immunity. *Nat Rev Immunol*. 2006;6(6):433–46.
- Lu R, Pan H, Shively JE. CEACAM1 negatively regulates IL-1β production in LPS activated neutrophils by recruiting SHP-1 to a SYK-TLR4-CEACAM1 complex. *PLoS Pathog*. 2012;8(4):e1002597.
- Hosomi S, Chen Z, Baker K, Chen L, Huang YH, Olszak T, Zeissig S, Wang JH, Mandelboim O, Beauchemin N, et al. CEACAM1 on activated NK cells inhibits NKG2D-mediated cytolytic function and signaling. *Eur J Immunol*. 2013;43(9):2473–83.
- Adler H, El-Gogo S, Guggemoos S, Zimmermann W, Beauchemin N, Kammerer R. Perturbation of lytic and latent gammaherpesvirus infection in the absence of the inhibitory receptor CEACAM1. *PLoS One*. 2009;4(7):e6317.
- Adler H, Steer B, Juskewitz E, Kammerer R. To the editor: Murine gammaherpesvirus 68 (MHV-68) escapes from NK-cell-mediated immune surveillance by a CEACAM1-mediated immune evasion mechanism. *Eur J Immunol*. 2014;44(8):2521–2.
- Beauchemin N, Arabzadeh A. Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. *Cancer Metastasis Rev*. 2013;32(3-4):643–71.
- Pils S, Gerrard DT, Meyer A, Hauck CR. CEACAM3: an innate immune receptor directed against human-restricted bacterial pathogens. *Int J Med Microbiol*. 2008;298(7-8):553–60.
- Schmitter T, Agerer F, Peterson L, Munzner P, Hauck CR. Granulocyte CEACAM3 is a phagocytic receptor of the innate immune system that mediates recognition and elimination of human-specific pathogens. *J Exp Med*. 2004;199(1):35–46.
- Heinrich A, Heyl KA, Klaile E, Muller MM, Klassert TE, Wiessner A, Fischer K, Schumann RR, Seifert U, Riesbeck K, et al. Moraxella catarrhalis induces CEACAM3-Syk-Card9-dependent activation of human granulocytes. *Cell Microbiol*. 2016;18(11):1570–82.
- Sarantis H, Gray-Owen SD. Defining the roles of human carcinoembryonic antigen-related cellular adhesion molecules during neutrophil responses to *Neisseria gonorrhoeae*. *Infect Immun*. 2012;80(1):345–58.
- Hill DJ, Toleman MA, Evans DJ, Villullas S, Van Alphen L, Virji M. The variable P5 proteins of typeable and non-typeable Haemophilus influenzae target human CEACAM1. *Mol Microbiol*. 2001;39(4):850–62.
- Hill DJ, Virji M. A novel cell-binding mechanism of Moraxella catarrhalis ubiquitous surface protein UspA: specific targeting of the N-domain of carcinoembryonic antigen-related cell adhesion molecules by UspA1. *Mol Microbiol*. 2003;48(1):117–29.
- Virji M, Watt SM, Barker S, Makepeace K, Doyonnas R. The N-domain of the human CD66a adhesion molecule is a target for Opa proteins of *Neisseria meningitidis* and *Neisseria gonorrhoeae*. *Mol Microbiol*. 1996;22(5):929–39.
- Sintsova A, Wong H, MacDonald KS, Kaul R, Virji M, Gray-Owen SD. Selection for a CEACAM receptor-specific binding phenotype during *Neisseria gonorrhoeae* infection of the human genital tract. *Infect Immun*. 2015;83(4):1372–83.
- Zid M, Drouin G. Gene conversions are under purifying selection in the carcinoembryonic antigen immunoglobulin gene families of primates. *Genomics*. 2013;102(4):301–9.
- Kammerer R, Popp T, Hartle S, Singer BB, Zimmermann W. Species-specific evolution of immune receptor tyrosine based activation motif-containing CEACAM1-related immune receptors in the dog. *BMC Evol Biol*. 2007;7:196.
- Kammerer R, Popp T, Singer BB, Schlender J, Zimmermann W. Identification of allelic variants of the bovine immune regulatory molecule CEACAM1 implies a pathogen-driven evolution. *Gene*. 2004;339:99–109.

Received: 19 May 2016 Accepted: 9 November 2016

Published online: 16 November 2016



24. Dveksler GS, Pensiero MN, Cardellicchio CB, Williams RK, Jiang GS, Holmes KV, Dieffenbach CW. Cloning of the mouse hepatitis virus (MHV) receptor: expression in human and hamster cell lines confers susceptibility to MHV. *J Virol*. 1991;65(12):6881–91.
25. Kammerer R, Ruttiger L, Riesenberger R, Schauble C, Krupar R, Kamp A, Sunami K, Eisenried A, Hennenberg M, Grunert F, et al. Loss of mammal-specific tectorial membrane component carcinoembryonic antigen cell adhesion molecule 16 (CEACAM16) leads to hearing impairment at low and high frequencies. *J Biol Chem*. 2012;287(26):21584–98.
26. Murata Y, Kotani T, Supriatna Y, Kitamura Y, Imada S, Kawahara K, Nishio M, Daniwijaya EW, Sadakata H, Kusakari S, et al. Protein tyrosine phosphatase SAP-1 protects against colitis through regulation of CEACAM20 in the intestinal epithelium. *Proc Natl Acad Sci U S A*. 2015;112(31):E4264–4271.
27. Zhang H, Eisenried A, Zimmermann W, Shively JE. Role of CEACAM1 and CEACAM20 in an in vitro model of prostate morphogenesis. *PLoS One*. 2013;8(1):e53359.
28. Chang CL, Semyonov J, Cheng PJ, Huang SY, Park JI, Tsai HJ, Lin CY, Grutzner F, Soong YK, Cai JJ, et al. Widespread divergence of the CEACAM/PSG genes in vertebrates and humans suggests sensitivity to selection. *PLoS One*. 2013;8(4):e61701.
29. Pavlopoulou A, Scroilas A. A comprehensive phylogenetic and structural analysis of the carcinoembryonic antigen (CEA) gene family. *Genome Biol Evol*. 2014;6(6):1314–26.
30. Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC. A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Mol Phylogenet Evol*. 2004;33(1):197–213.
31. Matsuda Y, Uno Y, Kondo M, Gilchrist MJ, Zorn AM, Rokhsar DS, Schmid M, Taira M. A New Nomenclature of *Xenopus laevis* Chromosomes Based on the Phylogenetic Relationship to *Silurana/Xenopus tropicalis*. *Cytogenet Genome Res*. 2015;145(3–4):187–91.
32. Tchoupa AK, Schuhmacher T, Hauck CR. Signaling by epithelial members of the CEACAM family - mucosal docking sites for pathogenic bacteria. *Cell Commun Signal*. 2014;12:27.
33. Fossum S, Saether PC, Vaage JT, Daws MR, Dissen E. Paired opposing leukocyte receptors recognizing rapidly evolving ligands are subject to homogenization of their ligand binding domains. *Immunogenetics*. 2011;63(12):809–20.
34. Barclay AN, Hatherley D. The counterbalance theory for evolution and function of paired receptors. *Immunity*. 2008;29(5):675–8.
35. Virji M, Evans D, Griffith J, Hill D, Serino L, Hadfield A, Watt SM. Carcinoembryonic antigens are targeted by diverse strains of typable and non-typable *Haemophilus influenzae*. *Mol Microbiol*. 2000;36(4):784–95.
36. Barrow AD, Trowsdale J. The extended human leukocyte receptor complex: diverse ways of modulating immune responses. *Immunol Rev*. 2008;224:98–123.
37. Sironi M, Cagliani R, Forni D, Clerici M. Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nat Rev Genet*. 2015;16(4):224–36.
38. Jack RS. Evolution of Immunity and Pathogens. *Results Probl Cell Differ*. 2015;57:1–20.
39. Semon M, Wolfe KH. Preferential subfunctionalization of slow-evolving genes after allopolyploidization in *Xenopus laevis*. *Proc Natl Acad Sci U S A*. 2008;105(24):8333–8.
40. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011;8(10):785–6.
41. Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 2013;30(12):2725–9.
42. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci*. 1992;8(3):275–82.
43. Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W. PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res*. 2000;10(4):577–86.
44. Jin L, Li Y, Chen CJ, Sherman MA, Le K, Shively JE. Direct interaction of tumor suppressor CEACAM1 with beta catenin: identification of key residues in the long cytoplasmic domain. *Exp Biol Med (Maywood)*. 2008;233(7):849–59.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

